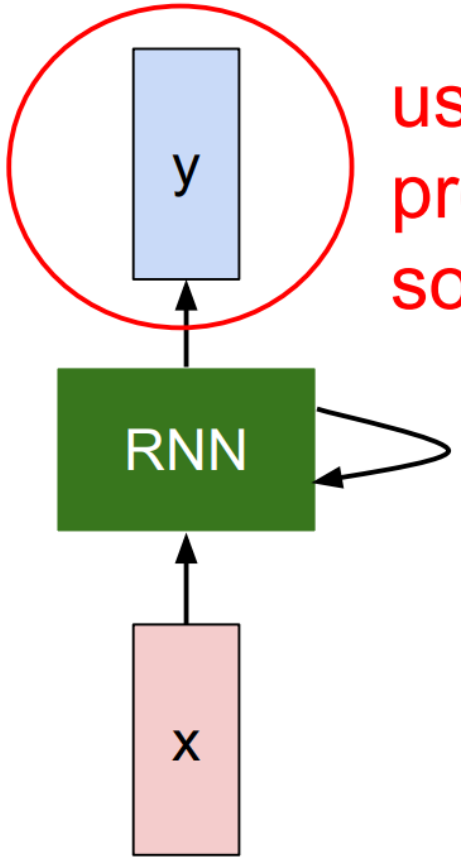
Artificial Intelligence

Long Short Term Memory Networks

Pham Viet Cuong
Dept. Control Engineering & Automation, FEEE
Ho Chi Minh City University of Technology

Recurrent Neural Network



usually want to predict a vector at some time steps

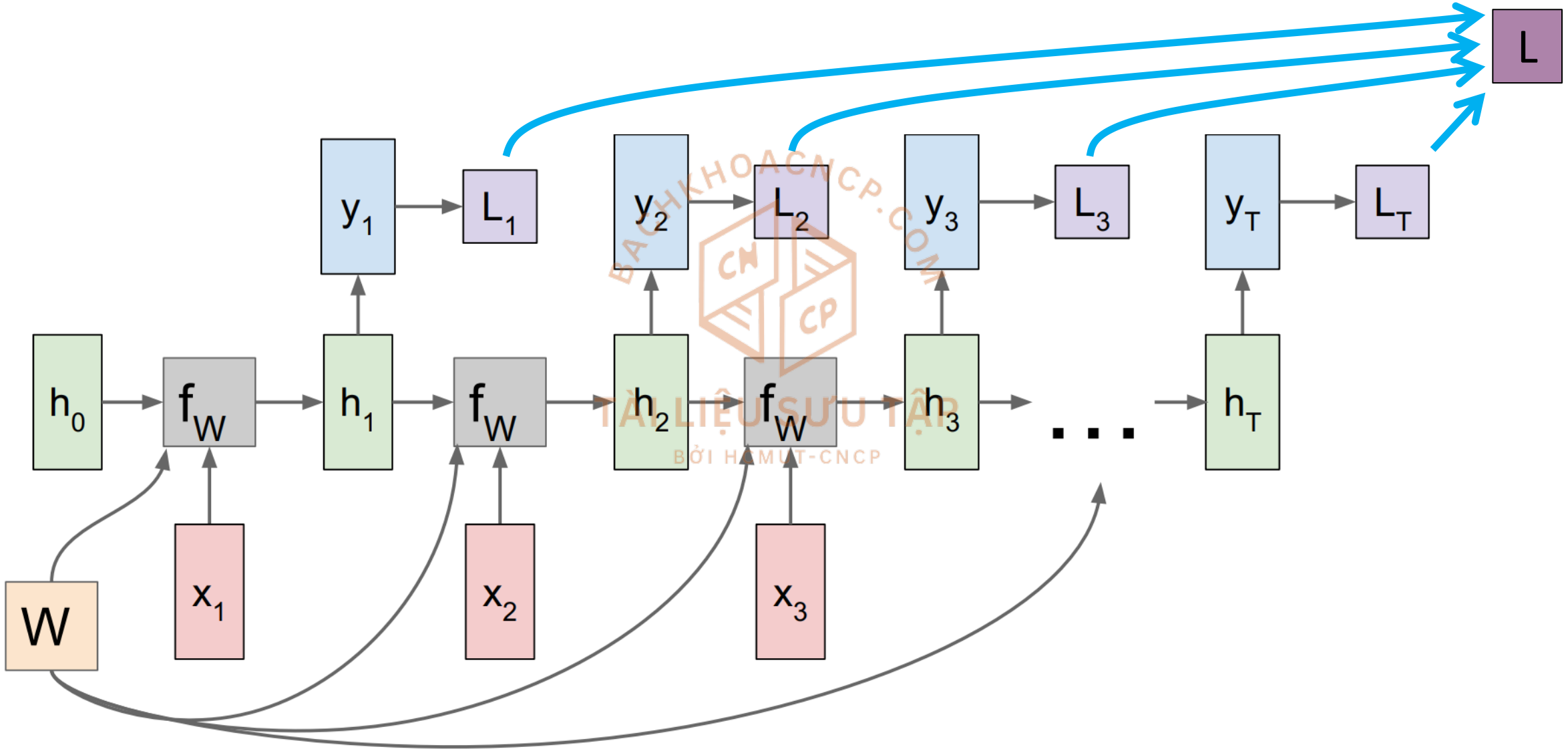
$$h_t = f_W(h_{t-1}, x_t)$$

new state — some function with parameters W — old state — input vector at some time step

$$h_t = tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$

$$y_t = softmax(W_{hy}h_t + b_y)$$

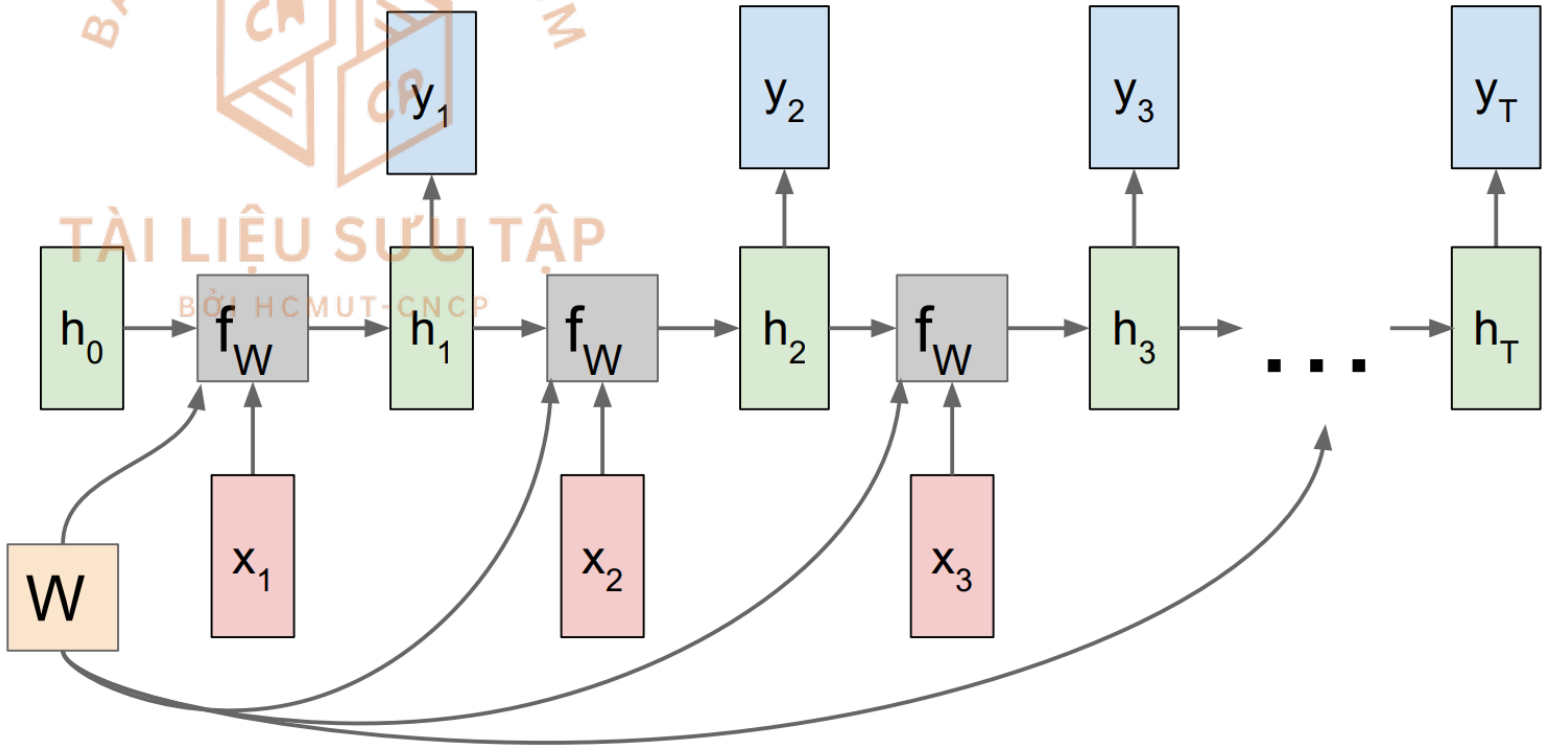
Recurrent Neural Network



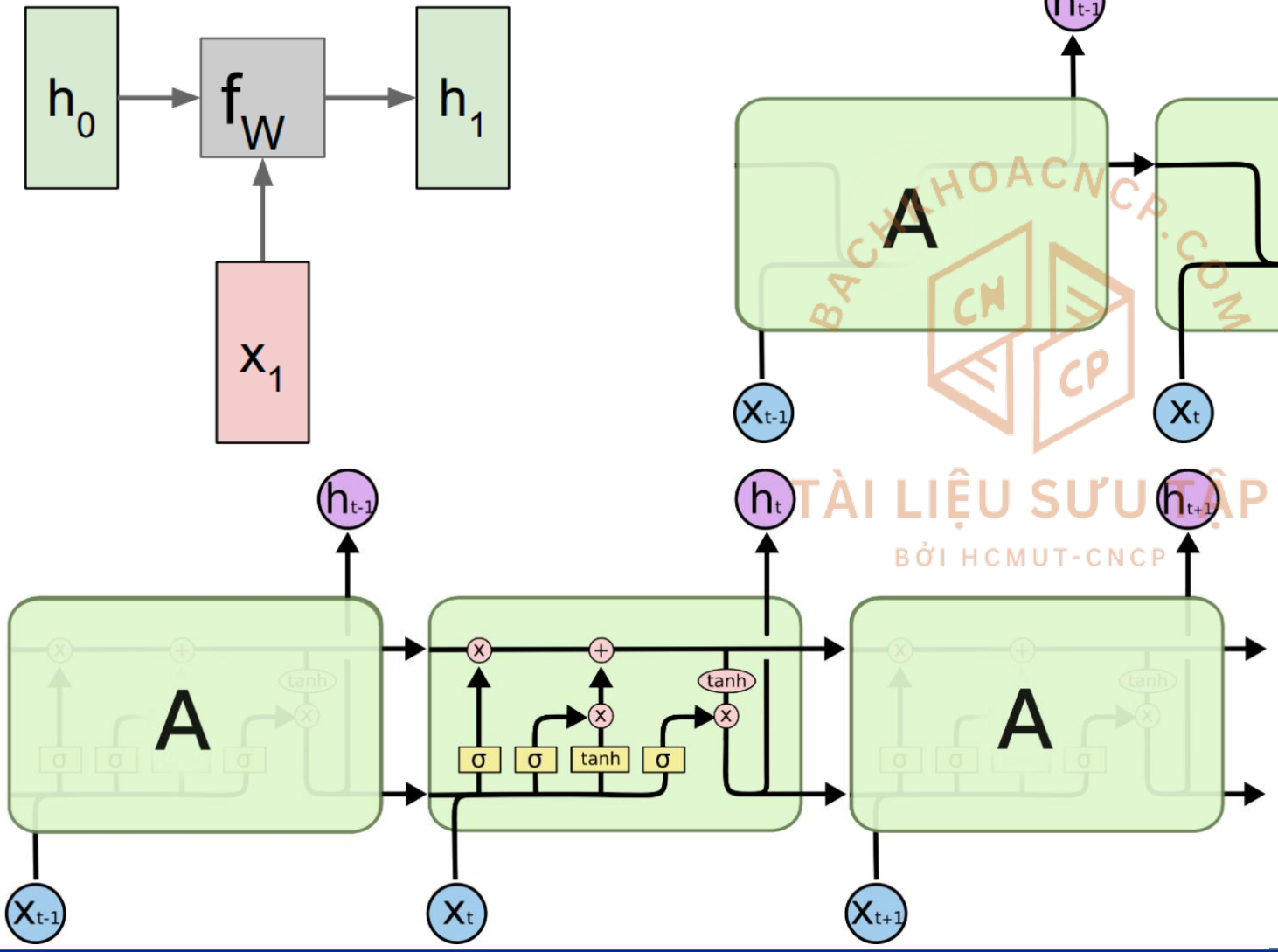
Long Short Term Memory

- RNN limitations:
  - Vanishing gradient
  - Exploding gradient
  - Suffering from long-term dependency

→ **LSTM**

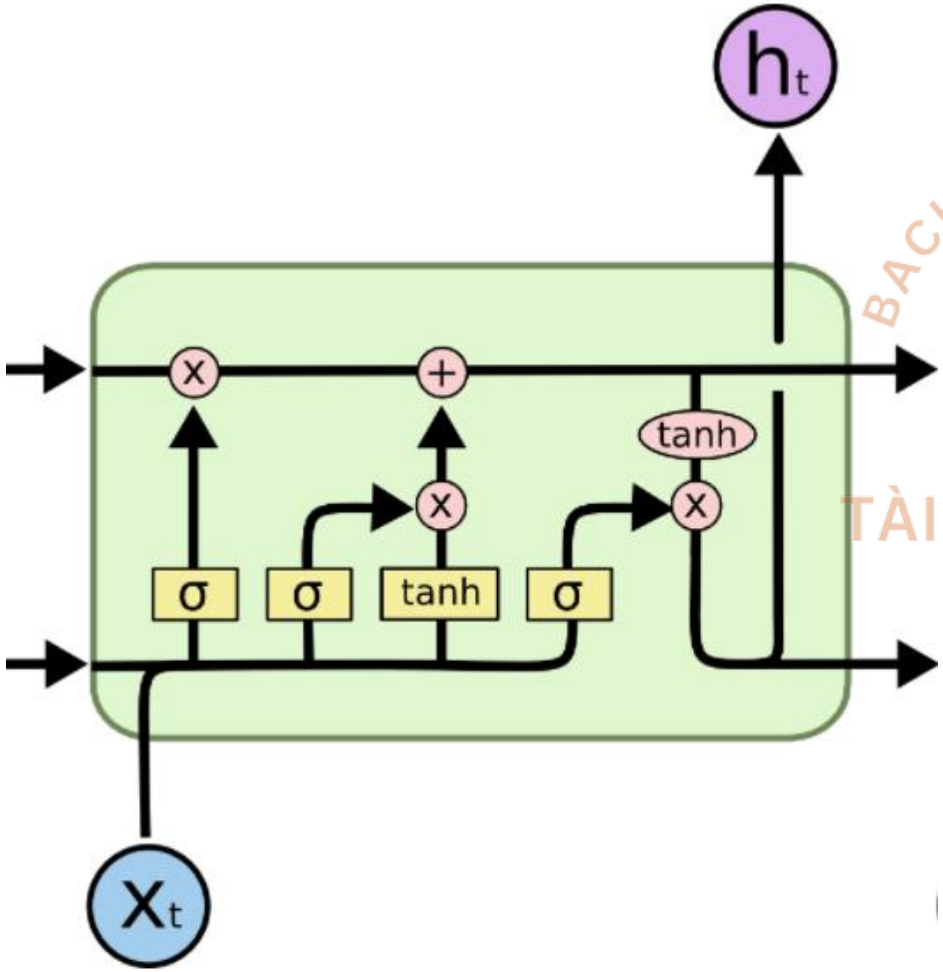


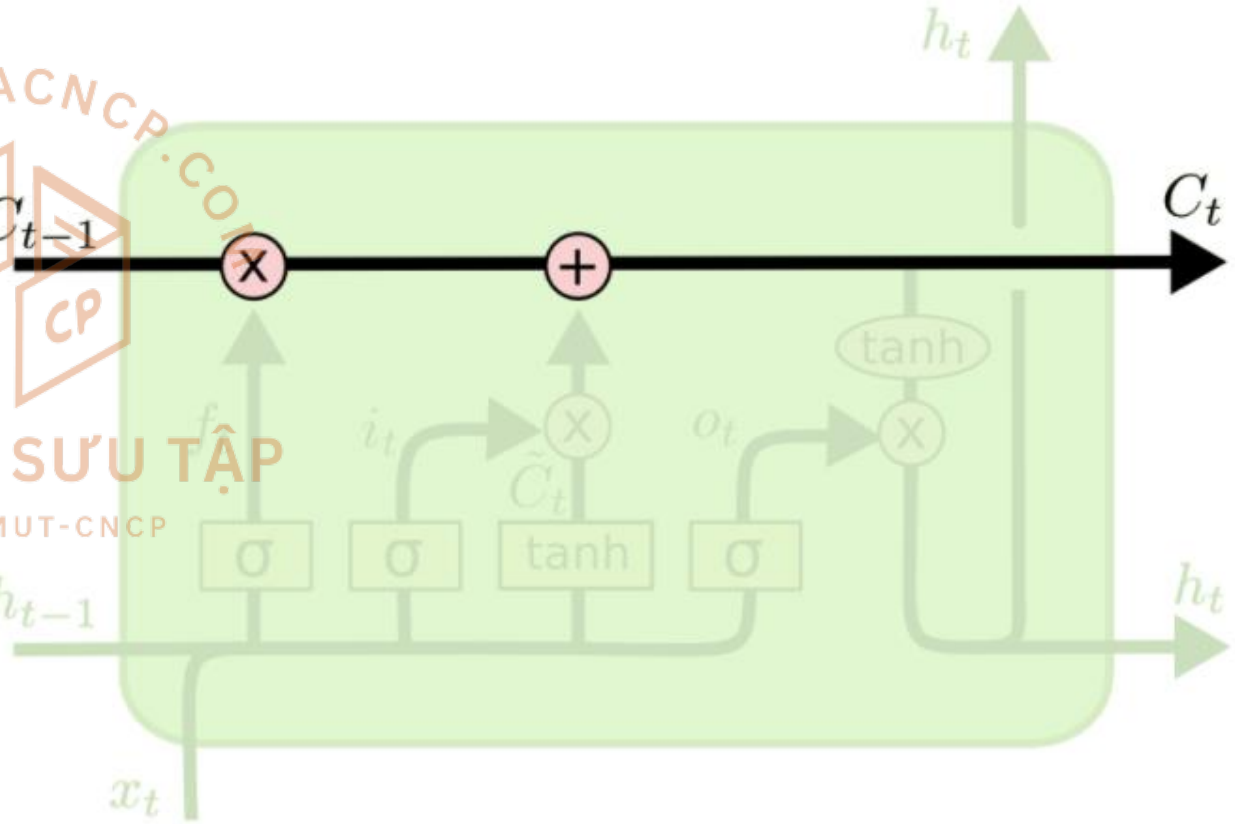
Long Short Term Memory



Long Short Term Memory

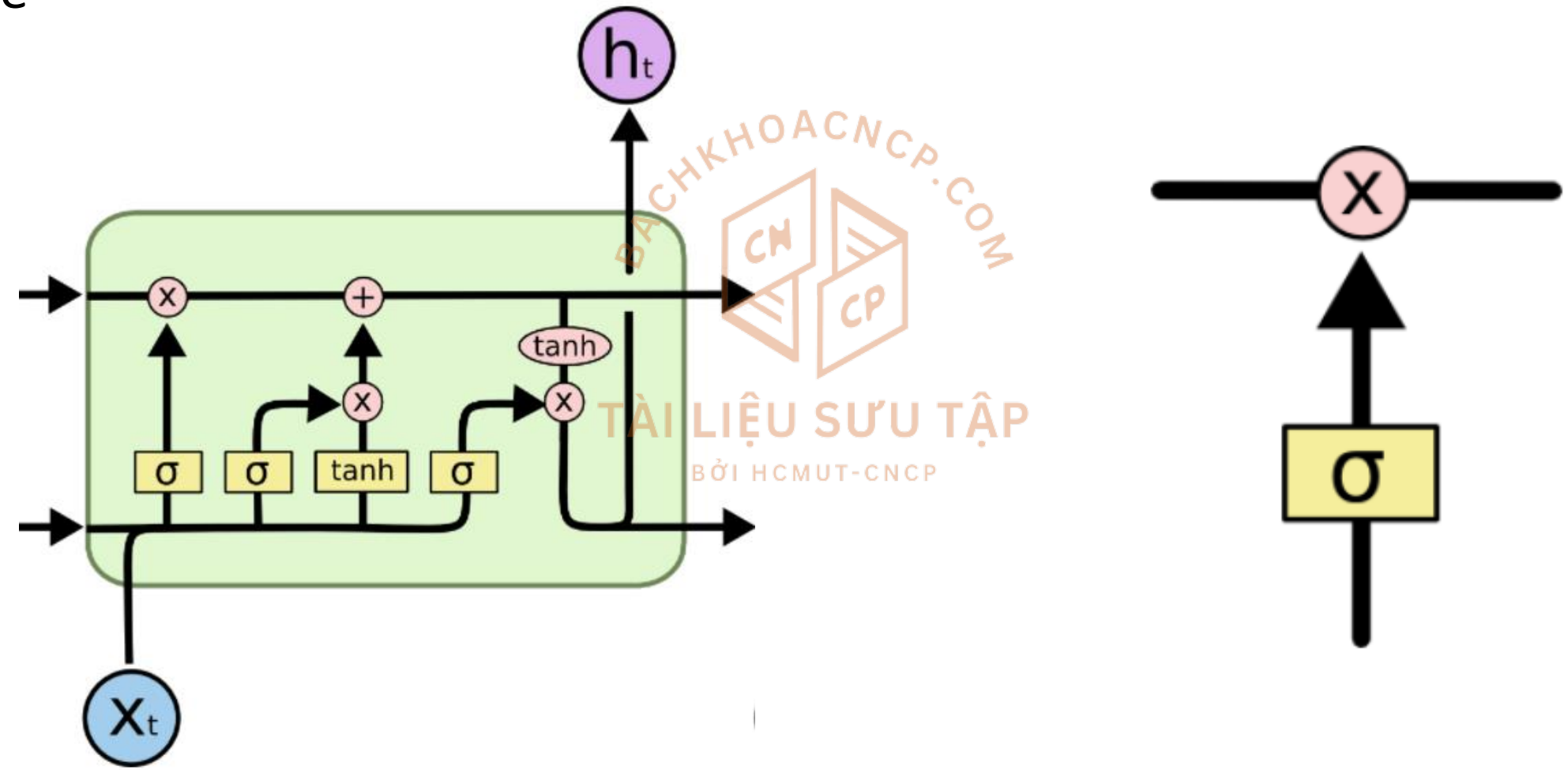
- ✓ Cell state





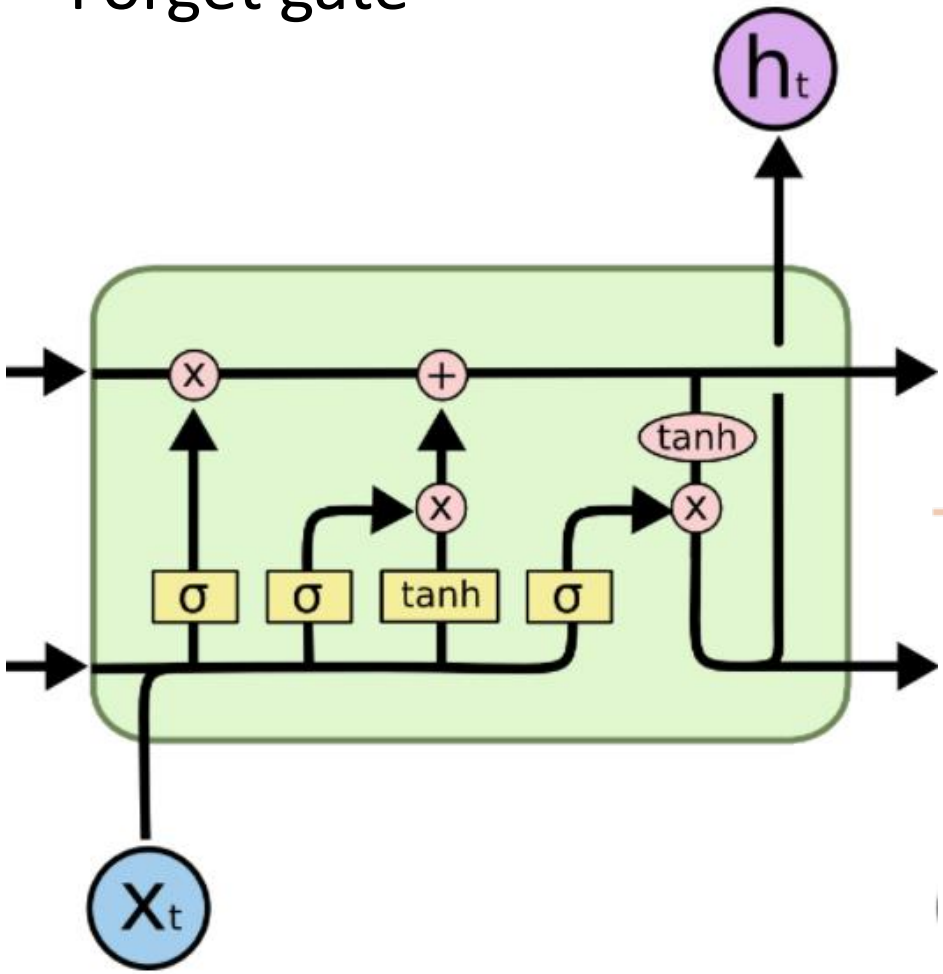
Long Short Term Memory

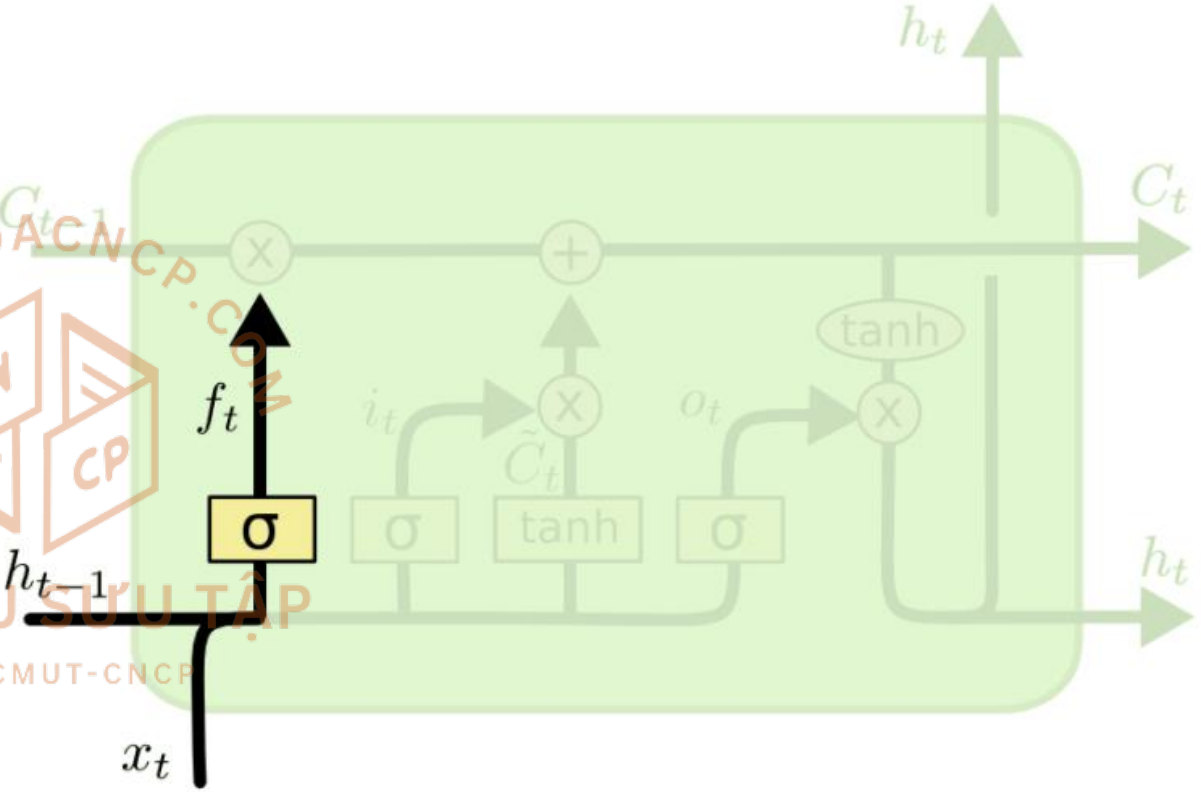
- ✓ Gate



Long Short Term Memory

✓ Forget gate

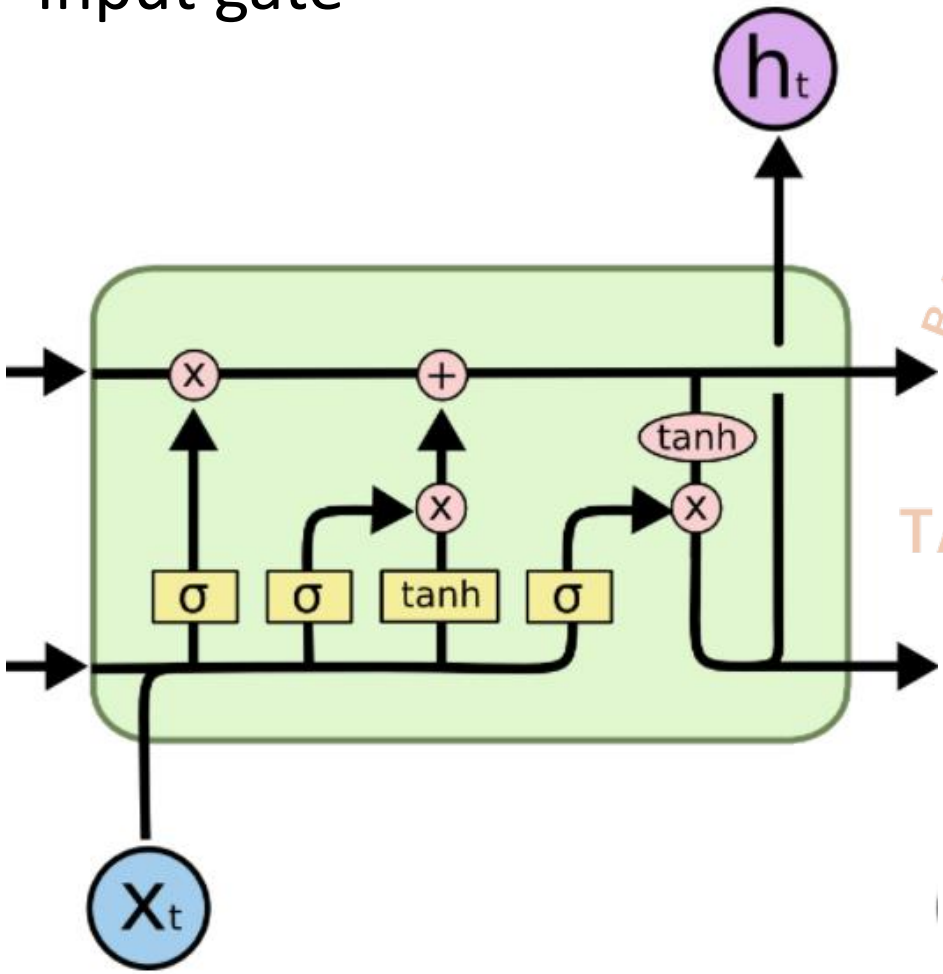


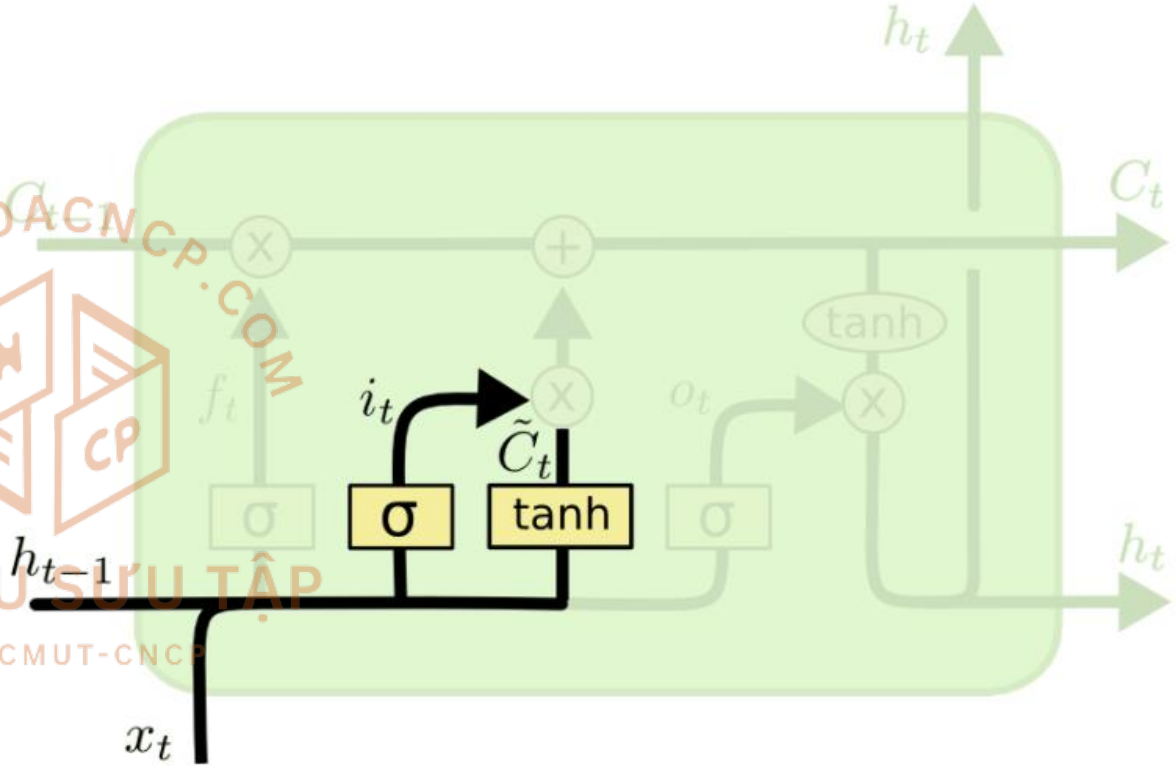


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Long Short Term Memory

- ✓ Input gate



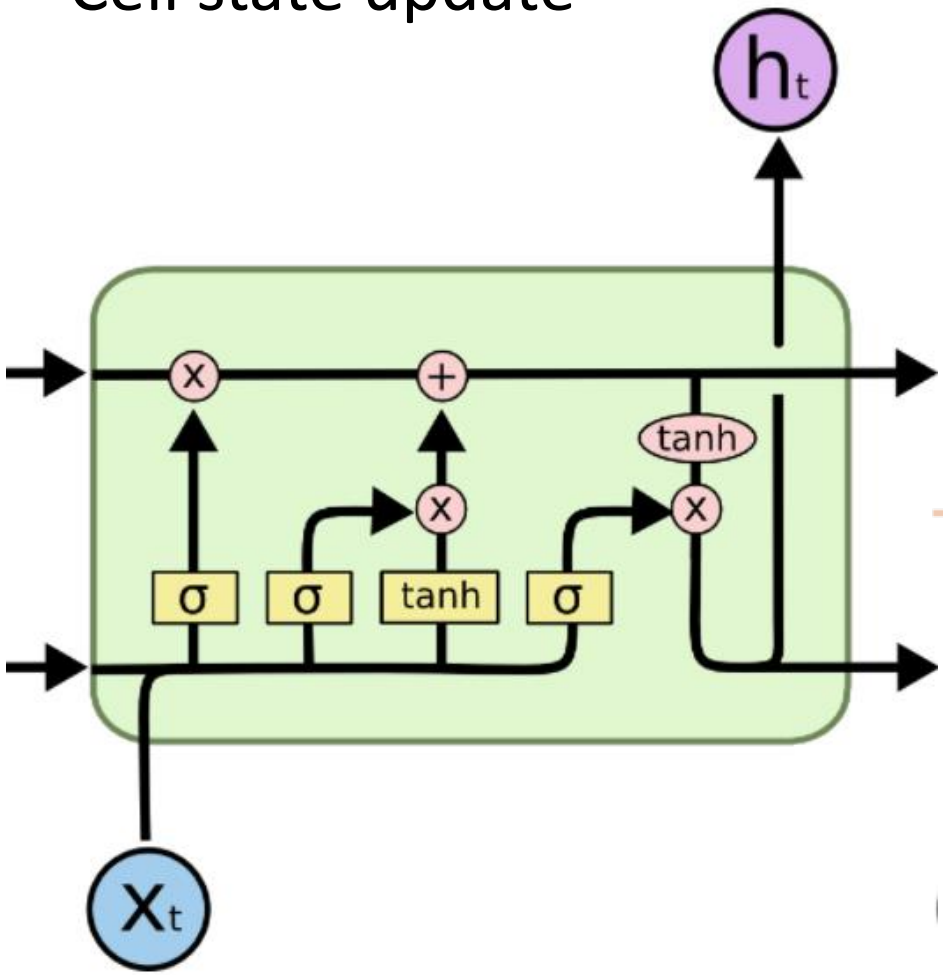


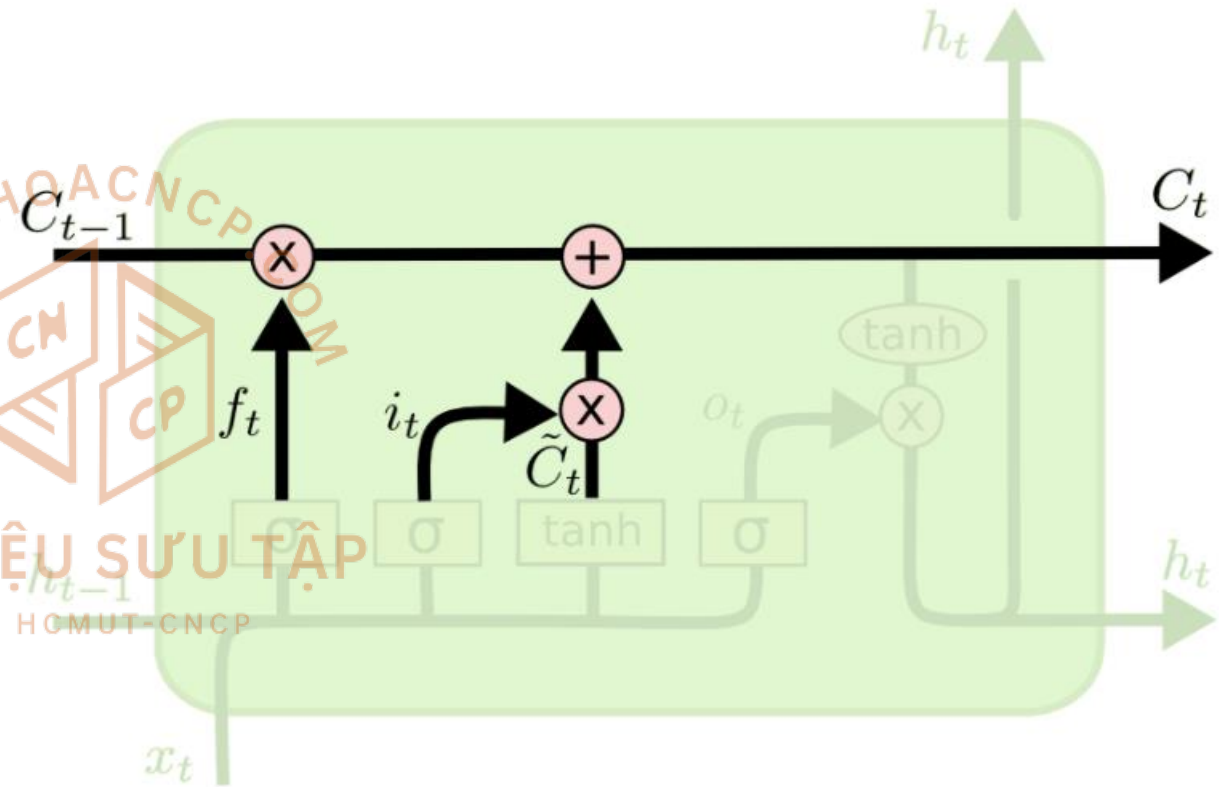
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Long Short Term Memory

✓ Cell state update

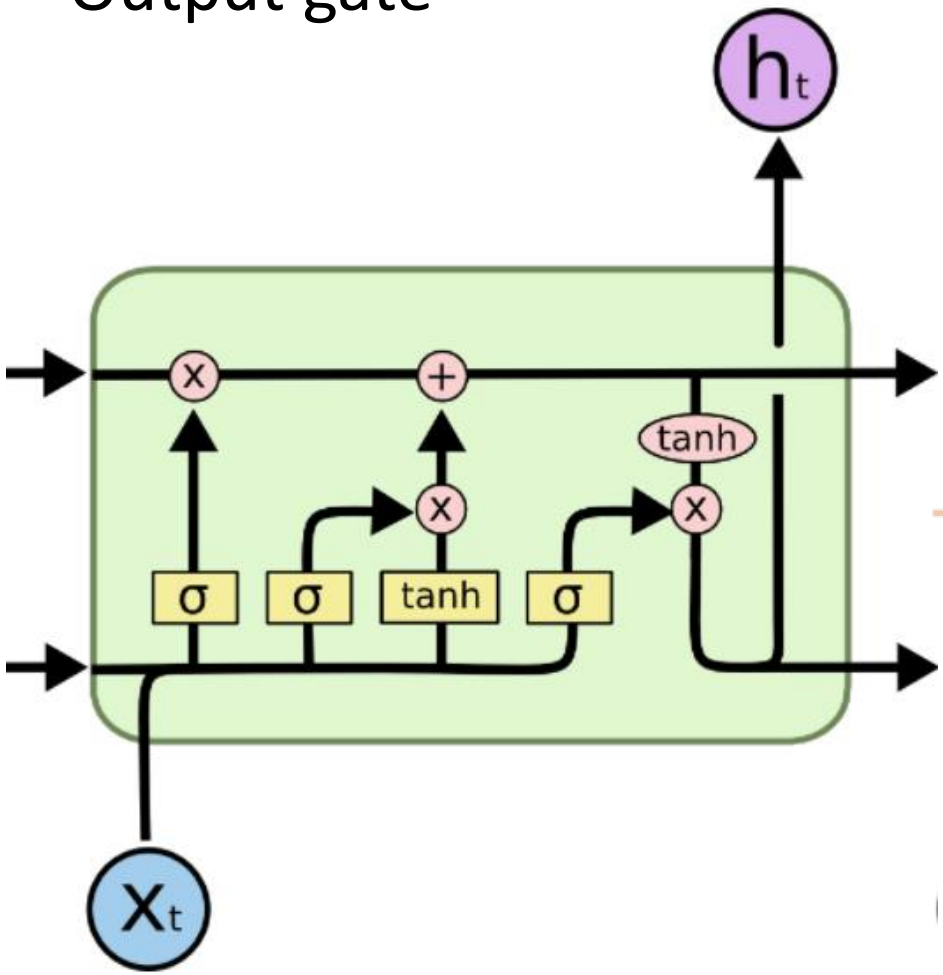


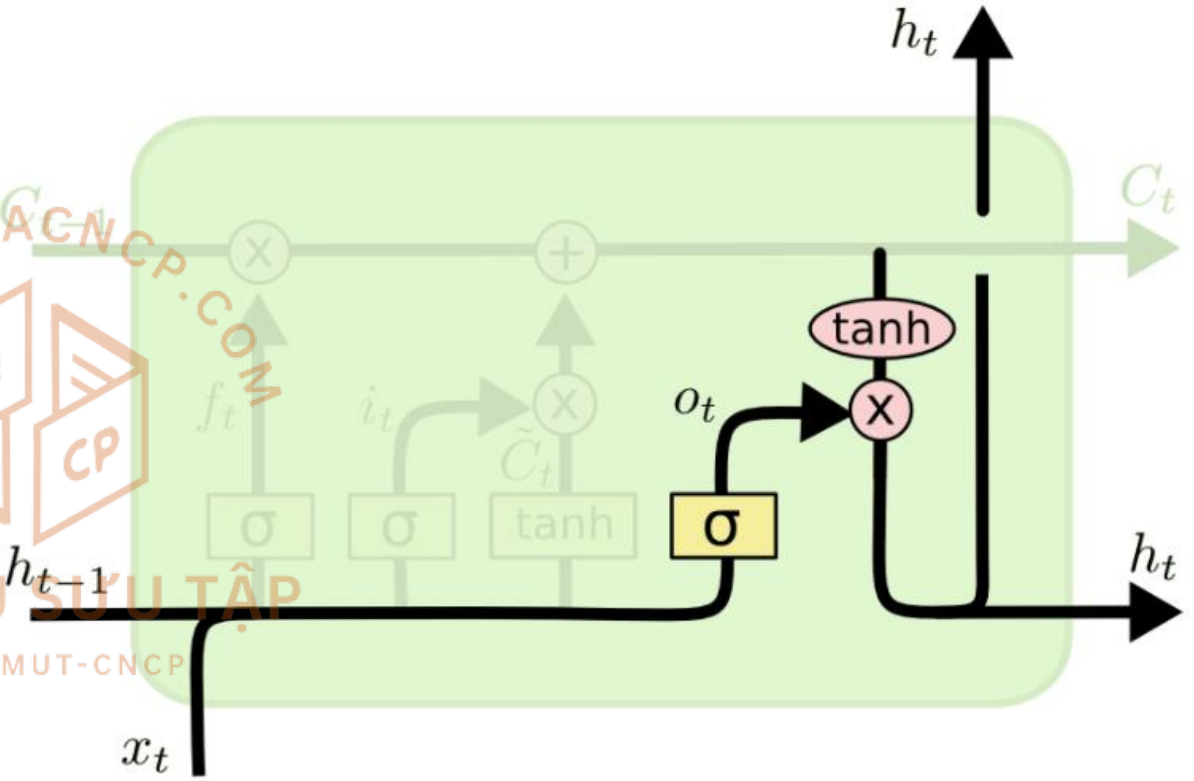


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Long Short Term Memory

✓ Output gate

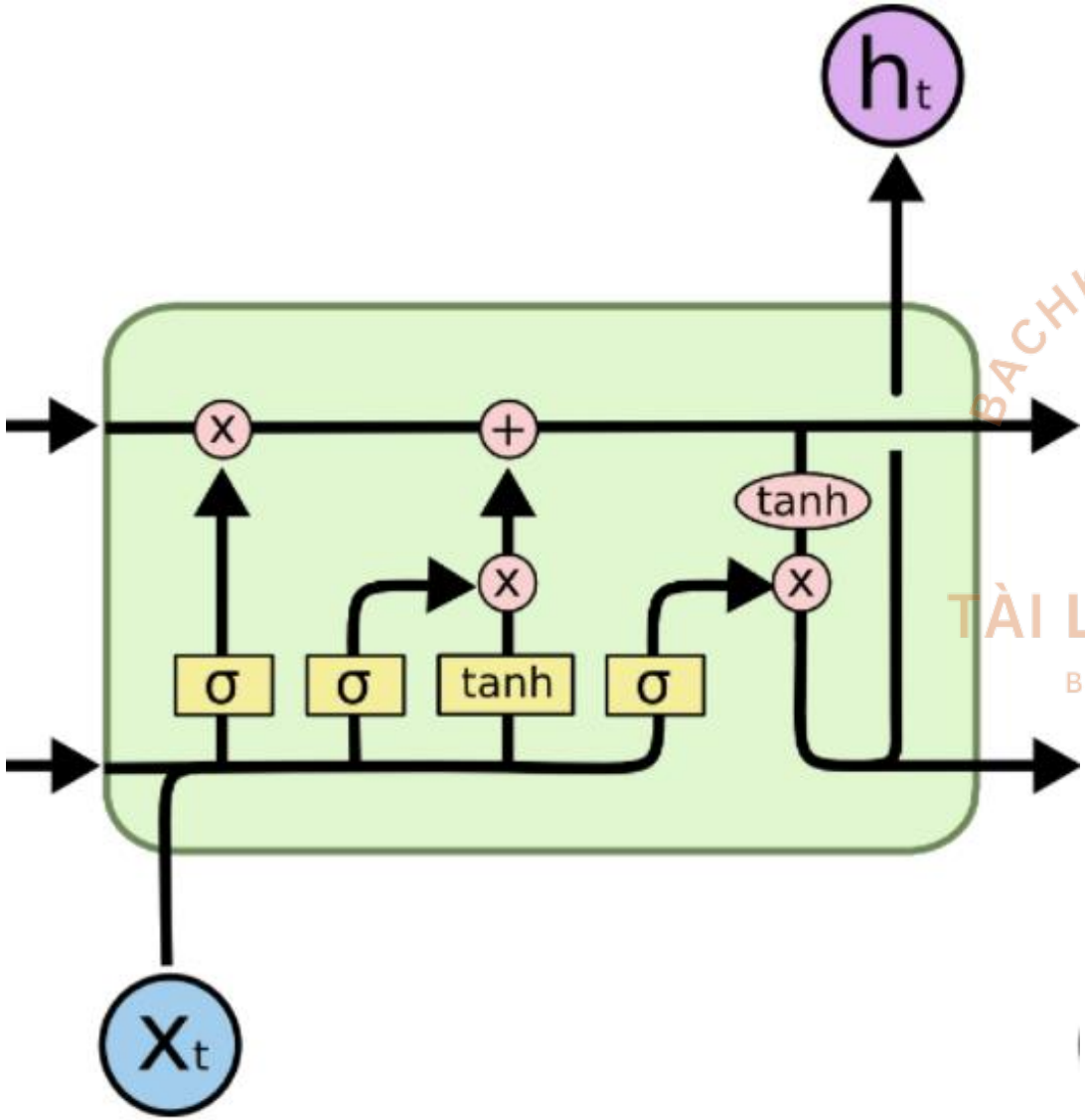




$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Long Short Term Memory



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

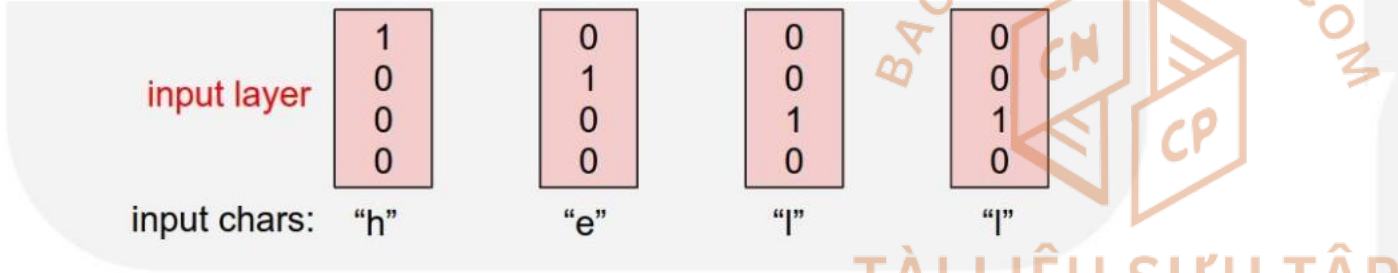
$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$

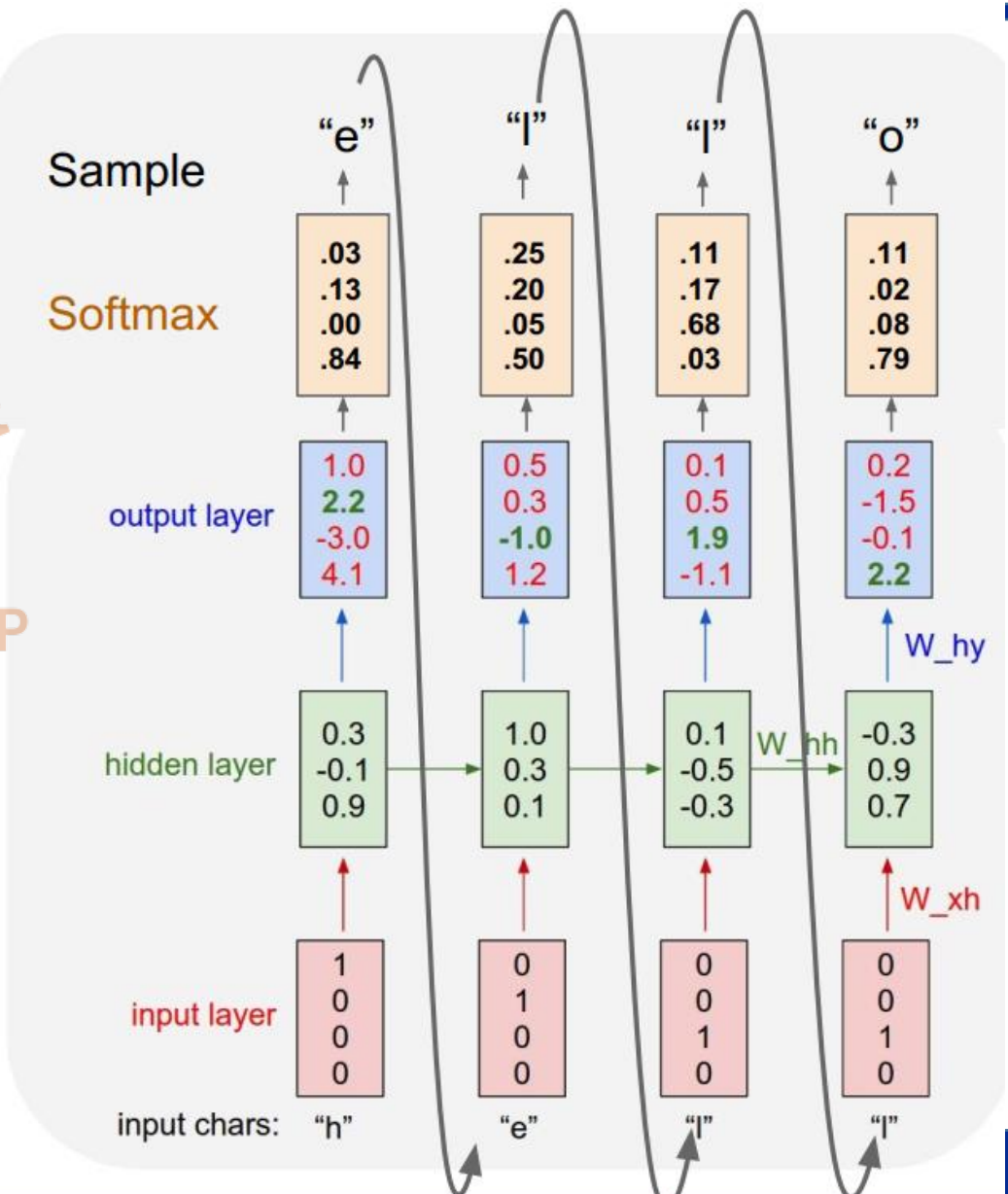
$$h_t = o_t * \tanh\left(C_t\right)$$

$$y_t = softmax\left(W_{hy} h_t + b_y\right)$$

Long Short Term Memory

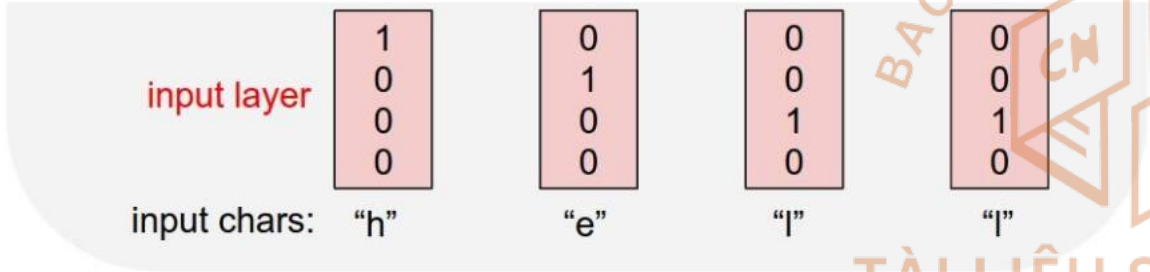
- Example: Language model - Character level
  - Training sequence: “hello”
  - Vocabulary: h, e, l, o





Long Short Term Memory

- Example: Language model - Character level
  - Training sequence: “hello”
  - Vocabulary: h, e, l, o



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$
$$y_t = softmax\left(W_{hy} h_t + b_y\right)$$

- x_t: 4x1 h_t: 3x1 C_t: 3x1
- W_f, W_i, W_C, W_o: 3x7
- b_f, b_i, b_C, b_o: 3x1
- W_{hy}: 4x3
- b_y: 4x1